

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
Before the Board of Patent Appeals and Interferences

Application No. : 09/348,425
Applicant : KIST, et al.
Filed : July 7, 1999
TC/A.U. : 2654
Examiner : Azad, Abul K.

Confirmation No. : 5214

#19
ART
12/10

RECEIVED

DEC 12 2003

Docket No. : 6169-125
IBM Docket No. : BOC9-1999-0036

Technology Center 2600

REPLY BRIEF UNDER 37 C.F.R. § 1.193

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Arlington, VA 22313-1450

Sir:

This Reply Brief is being filed pursuant to 37 C.F.R. § 1.193 in response to the Examiner's assertions and arguments set forth in the Examiner's Answer mailed on October 8, 2003 (the "Examiner's Answer"). This Reply brief is timely filed as a matter of right within the statutory two month period. Accordingly, no petition for extension of time under § 1.136(b) and no extension fee will be required. Notwithstanding, if the fees paid are deemed insufficient, authorization is hereby given to charge any deficiency to the undersigned's Deposit Account No. 50-0951.

Response to the Examiner's Response to Argument

I. Claim Language in Dispute:

As recited in the Appellants' independent claims 22 and 27, four fundamental limitations resonate throughout:

- (i) First, a voice command having two components, a voice command component and a dictation component, is identified within a contiguous utterance.
- (ii) Second, the voice command component is specified by a command grammar and the dictation component is free-form text which is not specified by the command grammar.
- (iii) Third, the dictation component is embedded within the voice command.
- (iv) Fourth, the identified voice command component is executed using at least a part of the dictation component as an execution parameter of the voice command.

II. Voice Command having Two Components

In the Examiner's Answer, it is asserted that the limitation of "identifying a voice command having a voice command component and a dictation component within a contiguous utterance" is taught by U.S. Patent No. 5,799,279 to Gould *et al.* (Gould). The Examiner has cited Figures 8a, 8b, 9a, and 9b, col. 1, line 55 – col. 2, line 13, and col. 5, line 13 – col. 6, line 67 in support of this assertion.

The Appellants' have explicitly claimed the step of identifying a voice command having a voice command component and a dictation component. Specifically, a single voice command is identified by the Appellants' invention that includes two separate and distinct components – a voice command component and a dictation component. Gould does not teach a voice command structure having two such components.

Instead, the cited portions of Gould illustrate only that Gould can process speech in parallel, with one path processing received speech as if it were a command and the other path processing speech as if it were dictated text. Gould then makes a determination as to what the recognized speech is – a command or dictated text. As such, the Gould invention "allows users to intermittently execute commands that affect the text (e.g., underlining or bolding particular words) without requiring the user to switch between separate command and dictation modes." (col. 2, lines 8-11) Gould does not, however, teach a voice command having two components – a voice command component and a dictation component.

In illustration, at col. 1, lines 44-54, Gould specifically states that:

The recognizing may include evaluating the likelihood that a given [speech] element is either a command element or a text element. The recognizing may be biased in favor of a given element being text or a command. The biasing may include determining if a given one of the elements reflects a command reject or conforms to a command template; or comparing recognition scores of the given element as a command or as text; or determining the length of silence between successive ones of the elements or whether the actions of the user imply that a given one of the elements cannot be text. (emphasis added)

At col. 1, line 55-57, Gould goes on to state that "recognizing may include, in parallel, recognizing the [speech] elements as if they were text, and recognizing the elements as if they were commands." These passages illustrate that Gould evaluates a portion of speech and makes a determination as to whether that speech is dictation or a command.

In cols. 5 and 6, Gould further illustrates that speech is recognized as either dictation or a command. For example, at col. 6, lines 14-29, Gould states:

While a user's speech is being recognized, the CPU sends keystrokes or scripting language to the application to cause the application to display partial results (i.e., recognized words within an utterance before the entire utterance has been considered) within the document being displayed on the display screen (or in a status window on the display screen). **If the CPU determines that the user's speech is text** and the partial results match the final results, then the CPU is finished. However, **if the CPU determines that the user's speech is text but that the partial results do not match the final results**, then the CPU sends keystrokes or scripting language to the application to correct the displayed text. **Similarly, if the CPU determines that the user's speech was a command**, then the CPU sends keystrokes or scripting language to the application to cause the application to delete the partial results from the screen and execute the command. (emphasis added)

From the above, it is clear that Gould describes a process where speech is processed and recognized as either a command or as dictated text.

While evaluating whether received speech is dictated text or a command, Gould presents initial recognition results on a display screen. This aspect was illustrated in the prior passage as well as at col. 6, lines 30-41, in describing Figures 8a, 8b, 9a, and 9b:

For example, the application being executed by the system is a meeting scheduler (FIGS. 8a, 8b, 9a, and 9b). After the system displays partial results 302 "schedule this meeting in room 507" (FIG. 8a), the system determines that the utterance was a command and removes the text from the display screen (FIG. 8b) and executes the command by scheduling 304 the meeting in room 507. Similarly, after the system displays partial results 304 "underline last three words" (FIG. 9a), the system determines that the utterance was a command and removes the text from the display screen (FIG. 9b) and executes the command by underlining 306 the last three words.

As noted, this passage illustrates that Gould displays initial recognition results prior to making a determination as to whether speech is a command or dictation. Pending the determination, Gould can correct those instances where a command was initially and

incorrectly recognized as dictated text. That is, Gould removes text that is initially displayed on a screen when that speech is later determined to be a command.

Thus, while Gould can distinguish between commands and dictated text when spoken one after the other, the portions of dictated text and commands are separate and distinct phrases that are not co-mingled or embedded within one another. As such, Gould does not teach a voice command construct having a voice command component and a dictation component.

III. Voice Command Component and the Dictation Component

In the Examiner's Answer, it is asserted that Gould teaches the claimed limitation of "wherein said voice command component is specified by a command grammar and said dictation component is free form text which is not specified by said command grammar." In support, the Examiner relies upon the same passages of Gould that were discussed in Section II.

In particular, within the Response to Argument section, the Examiner notes the example "schedule this meeting in room 507" from column 6, lines 30-41 of the Gould specification as teaching the identification of a voice command component and a dictation component. In columns 5 and 6, however, Gould discusses, at length, how commands are specified with particularity using templates. The discussion from columns 5 and 6 makes it clear that Gould uses a system or hierarchy of command vocabularies to specify the allowable words that may be spoken by a user and recognized as a command. Thus, the example "schedule this meeting in room 507", while appearing to be a mix of a command and dictation, is not. Rather, the entire

phrase is a command that is completely specified by the hierarchy of command grammars/vocabularies discussed by Gould. The command grammars enumerate each word that is allowable as a command as well as the ordering of those words.

Notably, as the example presented in Gould pertains to a scheduling application, that application has a command to reserve rooms for meetings. Thus, what appears to be dictation, specifically "in room 507", is actually part of the command to schedule meetings that is specified by a command template. As each word of this command is fully specified by a command grammar, it includes no dictation component.

In contrast, the Appellants' explicitly claim that the "voice command component is specified by a command grammar and said dictation component is free-form text which is not specified by said command grammar." As noted in the Appellants' Appeal Brief and in the Appellants' application, "ordinary dictation is a spoken utterance which does not contain a pattern of words recognizable by the system for controlling the operation of system or application software. Instead, dictation is spoken merely to have the system convert the spoken words into text within an electronic document." (pages 14, lines 7-10) "The dictation may be comprised of any set of words in a voice recognition vocabulary, which could consist of tens of thousands of words." (page 16, lines 3-4)

Comparing Gould to the Appellants' invention, the exemplary voice command discussed on page 19 of the Appellants' application, "load all files regarding first quarter results", has a command component "load all files regarding" and a dictation component "first quarter results". While the command component corresponds to a specified pattern of words, the dictation component, which is distinct from the command component, can include any words recognizable to the speech recognition system. In

other words, the dictation component "first quarter results" is not specifically enumerated or specified by a command grammar, but rather is composed of words that are generally recognizable to the speech recognition system.

Thus, while the Appellants' invention can identify a voice command having a dictation component, Gould requires that each word of a command be fully specified by a command grammar. In other words, Gould does not recognize voice commands having a dictation component. This is the case as Gould recognizes speech as either a command or as speech, but not as a command that can include a dictation component embedded therein.

IV. Dictation Component is Embedded within the Voice Command

While this limitation was not addressed in the Examiner's Response, from the above discussion and examples provided, it is apparent that Gould recognizes discrete portions of speech as either dictated text or as commands. By comparison, the Appellants' invention recognizes voice commands having both a command component and a voice component. As explicitly claimed, the dictation component is embedded within the voice command construct. Gould does not disclose such a feature.

V. Conclusion

The Appellants' have invented a system adapted for speech recognition that can identify voice commands having two components – a voice command component and a dictation component. While the voice command component is specified by a command grammar, the dictation component is free-form text that is not specified by the command

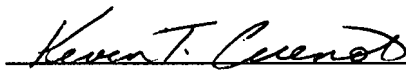
grammar. Rather, the dictation component can be composed of any word in a speech recognition vocabulary.

The limitations recited in the Appellants' independent claims require the identification of a voice command structure having both a voice command component and a dictation component, wherein the dictation component is embedded within the voice command. Gould, however, does not teach such a command structure. As such, Gould cannot be said to teach the system adapted for speech recognition of the present invention.

Accordingly, the Appellants believe that claims 22-31 are not anticipated by Gould under 35 U.S.C. § 102(e). It is thus submitted that the claims 23-31 define a patentably distinct invention over the prior art made of record, and a Notice of Allowance for claims 22-31 is accordingly and courteously solicited.

Respectfully submitted,

Date: 12/8/03



Gregory A. Nelson, Reg. No. 30,577
Kevin T. Cuenot, Reg. No. 46, 283
AKERMAN SENTERFITT
Post Office Box 3188
West Palm Beach, FL 33402-3188
(561) 653-5000